Public Comment for the May 22th meeting of the HIT Council

At the last meeting, the question was asked to explain normalized data. From the answer, I am concerned that the data produced might not lead to an accurate judging of the medical care system. The normalization process of the data inherently injects some distortions, as the data are taken from the electronic medical record and then put into categories and classified according to someone else's (Zato's) research, opinion and computer programs. Blood pressure numbers might be fairly straight forward for normalization, but the characterization of chest pain from a narrative might not be. It is important that the research data be precise and truthful as it will be used by the federal government and insurers to determine the optimal treatments for patients and which treatments will or will not be provided. Thus, providers might want to retain more direct control of what data are actually given to researchers.

Since aggregated study data may not apply to the complexities of an individual's body, great care must be given as to how research is converted into guidelines for patient care. The structure of the SIM program must not promote, but guard against a work environment where guidelines become gospel and where it will be safer for providers to follow guidelines (and some providers will not have enough training or breadth of knowledge to do otherwise) than to take the risk to work out of the box and explore what is best for the individual before them. How are "quality" care and expected outcomes going to be defined, particularly if the aggregated research data may not apply to an individual patient?

The Zato designers said that the data would not leave the control and site of the provider, but then how are the data collated with other provider data without sending data for reports over the internet and thus subject to breaches, hackers, etc.? Is it actually the case that that no patient data flows back to them through their system?

True patient privacy is crucial to the SIM program being successful. We would not want patients to not seek care or delay it until the treatment of their advanced illnesses costs the health care system even more. The State of CT is asking providers to encourage patients to reveal the history of their sexual and domestic abuse because it impacts on their health, but how fair is that to the patient when hundreds of authorized people will then be able to read that information in their record?

EHR systems need to be technically structured to address the following two problems:

First, patient right of consent was removed by HHS in 2002 to show identified records for treatment, payment and health care operations (oversight, quality control, tech support, business associates, covered entities, etc.), which means that there are hundreds of authorized people who can see any medical record. (Before the use of EHRs and the internet, this did not pose so much of a privacy problem as it does now.) So we need consent registries, segmentation for patients to keep some information from the full record and an indication for the provider that information was left out. (Also the technology must be developed to remove

at least the patient name and address from an EHR, during its perusal by oversight agencies, researchers, business people, etc.) Hopefully, SIM has not dropped this as a priority due to cost and expediency, because it will be much more costly and complicated to add these technologies later when they will be demanded by the public.

Second, 63-87% of people can be identified by using their gender, full date of birth and ZIP code (Latanya Sweeney, PhD). This rate of re-identification can be achieved by using that demographic data alone. The so called "HIPAA de-identified data" has had 18 identifiers removed but retains the year of birth and of all events of medical service, the gender and the 3-Digit ZIP code if the population locale is greater than 20,000. This data with the year, gender and 3 Digit ZIP *alone* has a .04% (2/5000 or 1440 people in our state of about 3.6 million) chance of being re-identified *without* adding the accompanying medical data. Thus the rate would be much higher if the individual's medical history narrative was added to those demographics and merged and cross referenced with the ever increasing available online data bases and would move toward the 63-87% rate re-identification rate.

Limited data sets are of particular concern as they have added back to them the full date of birth and ZIP code, making the data subject to the 63-87% re-identification rate. So if one has the medical information either from the APCD or the EHR, and even just gender, year of birth and 3-Digit ZIP code of "de-identified data," it's going to be possible to re-identify many more people that the .04% and closer to the 63-87% figure, that is many more than 1440 people in CT. Thus, there really is no such entity as safe "de-identified data" and patients must be asked for their consent before their intimate information is taken, even if someone else decides it is in their best interest or for the common good. Right now patients are forced to submit to losing control over who sees their health records in order to receive medical treatment and to buy health insurance. As yet, most people have no idea about the 2002 changes made by HHS, nor about the creation of the APCD that enable the handling of their data without their consent.

Thank you very much for your consideration of these issues.

Sincerely,

Susan Israel, MD